Data and Metadata Management at DIAS: Toward More Open Earth Environmental Information Platform

Toshiyuki Shimizu[1]*

[1]* *Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan*
Email: tshimizu@i.kyoto-u.ac.jp

Summary. Data Integration and Analysis System (DIAS) has continuously collected and managed earth observation data, and works as one of core research data platforms in Japan. As we have various kinds of data in the storage of DIAS, we are managing data by creating metadata in the dataset granularity. From 2016, phase III of DIAS started with more focus on open science, and we have started assigning Digital Object Identifiers (DOIs) to subset of datasets stored in the DIAS repository including the datasets we accepted from outside of DIAS. We are also focussing on metadata quality for better findability of datasets.

Keywords. DIAS, metadata, open science, metadata quality.

1. Introduction

Data Integration and Analysis System (DIAS) collects and stores data related to earth observation, and provides research platform for analysing data [1]. DIAS started from 2006, and after the successful completion of phase I and phase II of DIAS, the current project is in phase III which has started from 2016 with the aim of its practical operation.

Various kinds of datasets such as satellite data, ocean observation data, reanalysis data, land use data, etc. are stored in the DIAS repository, and many applications have been developed on them. Since we need to handle variety of datasets, we have developed the systems for managing metadata in the dataset granularity.

Also, with the movement of open science, the DIAS open science special interest group was launched, and we are discussing to make DIAS more open platform.

2. Data and Metadata Management

2.1 Data registration procedure

The requests of data deposit to DIAS are done with the submission of the application forms. After the review process, the data are stored in the DIAS repository if they are approved.

All datasets stored in the DIAS repository must have metadata, and data providers need to create metadata using our system.

2.2 Metadata management

The dataset metadata of DIAS is designed to be applicable to various kinds of datasets stored in the DIAS repository. It contains the fields for dataset name, contact point, creator, abstract text, topic category, spatiotemporal information, etc. The internal metadata format is ISO 19139, and each field is mapped to the corresponding element.

Once metadata is created using our system, the dataset becomes open to the public, and searchable from the DIAS dataset search and discovery system [2]. Users can specify the desired dataset and download data files through the system (Figure 1).

3. Open Science Activities

In our current project, phase III of DIAS, we focussed more on open science. One of the outcomes of the activities of the DIAS open science special interest group is assigning Digital Object Identifiers (DOIs) to datasets. The first assignment of a DOI in DIAS was done on March 2017 [3], and now we have several datasets with DOIs.

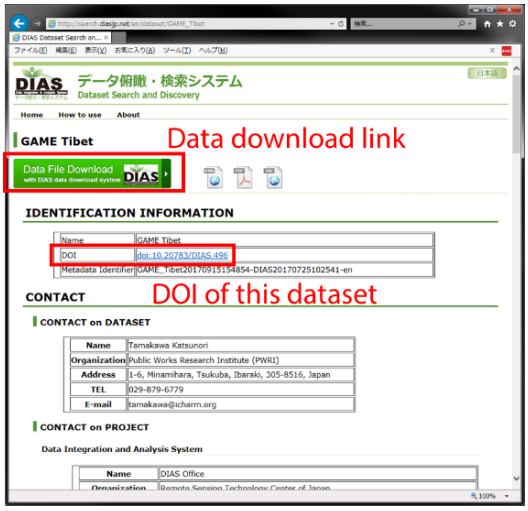

Figure 1. An example of a dataset metadata on the DIAS dataset search and discovery system [2].

The DOI of a dataset is stored in the metadata of the dataset, and we can find it at the dataset page in the DIAS dataset search and discovery system as shown in Figure 1.

Our other topics include getting official certificates of trustworthy data repositories so that DIAS can be considered as trustworthy from stakeholders.

4. Current and Future Prospects

4.1 DIAS as a national repository

Thus far, the stored datasets in the DIAS repository were mainly datasets from inside of DIAS or the collaborative projects. However, we have accepted some datasets outside of DIAS recently.

Since more and more journals request evidence data for papers, the demands for the data repositories are also increasing. However, data repositories which can store large volume of data are limited. Since DIAS has a very large storage space, it can be a candidate of a data repository especially for the valuable research data in Japan.

4.2 Focussing on metadata quality

For the end users of DIAS, findability of datasets stored in the DIAS repository are important, and thus we are focussing on metadata quality for providing better services.

As one of the attempts of improving metadata quality, we developed keyword recommendation method [4] since we considered keyword information is important for searching and categorizing datasets.

Besides, since we have not only metadata of DIAS datasets but also metadata from data centres outside of DIAS in the DIAS dataset search and discovery system, integrated management of different types of metadata is another topic concerning metadata quality.

5. Conclusions

DIAS is not only a data repository, but also an information platform for data science. We are managing various kinds of datasets through the metadata. About open science, the DIAS open science special interest group leads the discussion, and started assigning DOIs to datasets. We would like to make continuous efforts in order to make DIAS more open platform both for data providers and users.

Acknowledgments. I thank people in the DIAS open science special interest group, Dr. Asanobu Kitamoto, Dr. Masafumi Ono, Dr. Hiroko Kinutani, Dr. Masatoshi Yoshikawa, and Mrs. Yoko Nakahara for helpful discussion.

References

1. Kawasaki, A., Yamamoto, A., Koudelova, P., Acierto, R.A., Nemoto, T., Kitsuregawa, M., Koike, T., Data Integration and Analysis System (DIAS) Contributing to Climate Change Analysis and Disaster Risk Reduction. *Data Science Journal*, 16, 41, 1–17, 2017
2. DIAS dataset search and discovery system, http://search.diasjp.net/en [accessed on: Oct. 2017]
3. DIAS First DOI Registration, http://www.diasjp.net/infomation/press-release-dias-first-doi-registration/ [accessed on: Oct. 2017] (in Japanese)
4. Ishida, Y., Shimizu, T., Yoshikawa, M, A Keyword Recommendation Method Using CorKeD Words and Its Application to Earth Science Data. *11th Asia Information Retrieval Societies Conference*, 96–108, 2015